Journal of Multimedia Information System VOL. 5, NO. 2, June 2018 (pp. 111-114): ISSN 2383-7632(Online)
http://dx.doi.org/10.9717/JMIS.2018.5.2.111

Brief Paper:
Analysis of Traffic Accident using Association Rule Model

Sun-Young Ihm[1],Young-Ho Park[2*]

Abstract: Traffic accident analysis is important to reduce the occurrence of the accidents. In this paper, we analyze the traffic accident with Apriori algorithm to find out an association rule of traffic accident in Korea. We first design the traffic accident analysis model, and then collect the traffic accidents data. We preprocessed the collected data and derived some new variables and attributes for analyzing. Next, we analyze based on statistical method and Apriori algorithm. The result shows that many large-scale accident has occurred by vans in daytime. Medium-scale accident has occurred more in day than nighttime, and by cars more than vans. Small-scale accident has occurred more in night time than day time, however, the numbers were similar. Also, car-human accident is more occurred than car-car accident in small-scale accident.

Key Words: Traffic Accident, Association Rule, Apriori Algorithm, Data Mining.

I. INTRODUCTION

Korea has frequent traffic accidents because of its geographical and sociological characteristics compared to other countries. Traffic accident analysis is necessary to understand the factors affecting traffic accidents, and to reduce or control traffic accidents. Traffic accidents are caused by a combination of factors such as human factors, social factors, and environmental factors. Drinking, drowsiness driving, and driving carelessness are the main causes of traffic accidents, but there is not much research to analyze the association of various factors. Therefore, it is very interesting to understand the reasons and factors of traffic accidents in Korea.

In this paper, we find out the association rules based on various factors to analyze the causes of traffic accidents. We first collect the traffic accident data generated in Korea and preform the preprocessing phase for analysis. And then, we analyze the data to derive its meaning and discuss the results. Here, the analysis performs a basic statistical analysis considering various properties and we use the Apriori algorithm for the association analysis.

The rest of this paper is organized as follows. Section 2 describes the previous work related to this paper. Section 3 explains a model for traffic accident factor analysis, and Section 4 analyzes the cause of traffic accident accordingly. Section 5 summarizes and concludes the paper.

II. RELATED WORD

In this section, we discuss the related work. First, there have been various studies to analyze the cause and severity of traffic accidents. Li et al. [5] applied data mining algorithms such as Apriori algorithm, Naïve Bayses classifier and k-means clustering algorithm on the roadway traffic data. The authors performed the analysis, including weather, light conditions or surface condition. However, an analysis between more closely related attributes to the accident, such as the accident type or the road type is needed. Chong et al. [3] applied machine learning algorithms to model the severity of injury that occurred during traffic accidents. They considered neural networks trained using hybrid learning approaches, support vector machines, decision trees, and concurrent hybrid model involving decision trees and neural networks. Also, decision tree used for analyzing and founding the traffic accident patterns of N5 National Highway in Bangladesh in [8] Oña et al. [6] presented an analysis accidents on rural highway in Spain using Bayesian Networks to classify traffic accidents according to their injury severity. In [10], a model was proposed to predict the severity of traffic accidents in Abu Dhabi. Abellán et al. [1] proposed an effective method for extracting rules from decision tree to extract some important relationships between variables. And they applied the proposed method to obtain relevant rules from rural road traffic accident data in Spain. In [7], authors analyzed the traffic accident injury severity on Slovenian roads with a classification

Manuscript received March 10, 2018 ; Revised March 27, 2018 ; Accepted April 04, 2018. (ID No. JMIS-2018-0020)
Corresponding Author (*): Young-Ho Park, Sookmyung Women's University, 82-2-2077-7297, yhpark@sm.ac.kr.
[1]Big Data Using Research Center, Sookmyuung Women's University, Seoul, Korea, sunnyihm@sm.ac.kr
[2]IT Engineering, Sookmyung Women's University, Seoul, Korea, yhpark@sm.ac.kr

and regression tree algorithm. Castro and Kim [2] use Bayesian network, decision tree and artificial neural networks to detect factors with the greatest influence on car accidents in United Kingdom.

Also, there are several studies that analyze about traffic accidents in Korea. A structural equation model (SEM) [4] is proposed to correlate the variables related to traffic accidents. It uses accident data occurred on highways in Korea and estimate relationship among exogenous factors and traffic accident size. Sohn et al. [9] applied data mining techniques to estimate and classify severity types of road traffic accidents in Korea. They used neural network, logistic regression and decision tree to select a set of influential factors and to build up classification modes for accident severity.

III. BIG DATA MODEL DESIGN FOR TRAFFIC ACCIDENT FACTOR ANALYSIS

In this section, we design a big data model for traffic accident analysis and explain step by step. Figure 1 shows a model designed to analyze traffic accident factors. The first step is to define the problem and clear strategy for analyzing. In this paper, we define the problem as analyzing the causes and association rules of traffic accidents in Korea. The second step is to collect data for the big data analysis. We use traffic deaths data occurred in 2015-2016 in Korea for this study. Next, in the third step, we determine the properties of collected data and perform preprocessing. In the fourth step, we analyze the data with statistical method and Apriori algorithm. In the final step, we derive the insights based on the analyzed results.

IV. TRAFFIC ACCIDENT FACTOR ANALYSIS

In this section, we analyze traffic accident factors through various perspectives according to the model designed in Section 3.

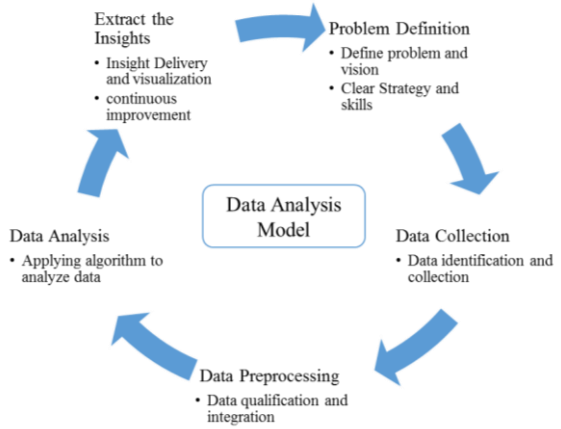

Fig. 1. A data analysis model.

1. Data Cleaning

In this section, we describe the data cleaning phase. First, we describe the data used in the analysis, and explain the preprocessing step of the collected data.

1.1. Data Collection

The aim of this study is to analyze the factors of traffic accidents in Korea. We use the traffic accident data provided by the government for the analysis, which are occurred in 2015 and 2016. The collected data includes 23 attributes, such as the date and time of the accident, type of accident, the number of deaths and injured people, etc., and the data is provided in csv format. Table 1 summarizes the main attributes of the collected data.

Table 1. Main attributes of the collected data

Attribute	Description
The date and time of accident	The year, month, day and time of the accident
Time slot	Time of accident (day or night)
Day of the week	Days of accidents (Mon., Tues., …, Sun.)
Number of deaths	The number of people who died due to traffic accidents
Number of seriously injured people	The number of injured people (More than 3 weeks of treatment is required)
Number of slightly injured people	The number of injured people (More than 5 days and less than 3 weeks of treatment is required)
Number of injured people	The number of injured people (Less than 5 days of treatment is required)
Area	Traffic accident area (city, country, district)
Type of accident	Types of accidents (Car-Car, Car-Human accidents, etc.)
Type of accident: subcategory	Type of behavior at the time of the accident (frontal collision, side collision, walking, etc.)
Road type	Type of accident road (highway, underpass, etc.)
Accident vehicle	Type of accident vehicle of the biggest fault or the smallest damage among the traffic accident (Truck, car, etc.)
Victim vehicle	Type of accident vehicle of the fewest fault or the biggest damage among the traffic accident (car, passenger, etc.)

[1] https://www.data.go.kr/dataset/15003493/fileData.do

1.2. Data Preprocessing

In this section, we describe the process of preprocessing the collected data for analysis. Since the data currently available only provide information relevant to the occurrence of a traffic accident, we could not determine the scale of the accident. Therefore, in this paper, we create a variable called ‘severity’ which can measure the scale of an accident, so that the scale of the accident can be quantified and the degree of accident ca be easily identified. The variable ‘severity’ is used to measure the scale of accident. For this, we weighted the variables

'Number of deaths', 'Number of seriously injured people', 'Number of slightly injured people' and 'Number of injured people', which indicate the number of traffic accident victims. And the severity of the accident is shown in Equation (1).

$$Severity = (Number\ of\ deaths \times 5) + (Number\ of\ seriously\ injured\ people \times 3) + (Number\ of\ slightly\ injured\ people \times 2) + (Number\ of\ injured\ people \times 1) \quad (1)$$

In addition, we create a variable 'Scale of accident' by categorizing the 'severity' to better understand. If the severity is higher than 50, it derives the attribute 'Large-scale accident'; if it is higher than 20, it derives 'Medium-scale accident'; if it is lower than 20, it derives 'Small-scale accident'.

2. The Result of Data Analysis

In this section, we analyze the data and describe the result. First, we performed statistical analysis based on the attributes of the preprocessed data. Figure 2 shows the result of analysis of traffic accident with type of accident, and the most accident occurred during the crossing.

Table 2 shows the result of traffic accident analysis according to the 'Scale of accident', which is derived from section 4.1.2. In Table 2, small-scale accident is occurred more than large-scale accident. Large-scale accidents are the least number of accident, however, they should be more cautious because the number of deaths and injured people. Also, it is considered the main reason of traffic accident varies depending on the scale of accident. Thus, in this paper, we analyze the traffic accident according to the scale of accident.

Table 2. The result of traffic accident analysis by 'Scale of accident'.

	Small-scale Accident	Medium-scale Accident	Large-scale Accident
Number of cases	6,738	103	15

Next, we analyze the traffic accident by scale of accidents. In this paper, we used Apriori algorithm, which is an algorithm to find out association rules to analysis.

2.1. Large-scale Accident Analysis

In this section, we analyze the large-scale accident which has severity higher than 50. The scale of this accident is large and it does not occurs often. For the experiment, we set the minimum support as 0.4 and the reliability as 0.8. The more frequent associate rules indicate that the time of accident is during the day time and the type of victim vehicle is the vans. It means that many large-scale accident occurs during the day time because the amount of vehicle traffic in day time is more than night time, and also there are many passengers on the vans. And many large-scale accidents occurs when the type of accident is 'ongoing-collision'. Also, if people do not follow the 'Safety driving obligation', there are many large-scale accident occurs.

2.2. Medium-scale Accident Analysis

In this section, we analyze the medium-scale accident which has severity higher than 20 and lower than 50. First, medium-scale accident has more occurred during day time than night time. However, it is more occurred during day time than night time only 1.16 times, so there is no big difference. This means that medium-scale accident occurs frequently regardless of time of day. In addition, both vehicle of the accident and victim are 'passenger cars', and it means the number of passengers are less than that of 'van', but the number of accident is higher.

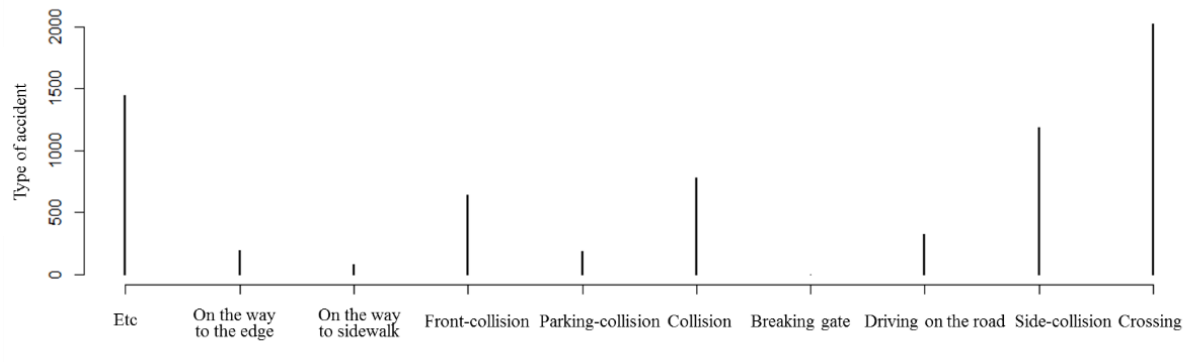

Fig. 2. Traffic accident analysis by type of accident.

2.3. Small-scale Accident Analysis

In this section, we analyze the small-scale accident which has severity lower than 20. Small-scale accident are most frequent because the number of death and injured people is lower than other accidents. Unusual point in small-scale accident is that it more occurs during night time than day time. Though it is a slight difference, it means small-scale accident frequently occurs during night time whereas other accidents are more occur in day time. Also, there are the victim is most likely a 'pedestrian' rather than a vehicle. This means that in the case of a small-scale accident, there are many accidents occur between vehicle and pedestrians rather than between vehicles. Table 3 shows some of the results to find out an association rules from large-scale accident data. The rule that has higher support and confidence is the more associated rule.

Table 3. The result of traffic accident analysis

lhs	rhs	support	confidence
{daytime}	{Large-scale accident}	0.933	1.00
{Accident vehicle: Van}	{Large-scale accident}	0.600	1.00
{Driving on the road}	{Large-scale accident}	0.533	1.00
{daytime, do not following the *Safety driving obligation*}	{Large-scale accident}	0.533	1.00
{Cross road}	{Large-scale accident}	0.400	1.00
{daytime, Victim vehicle: Passenger car}	{Large-scale accident}	0.400	0.857

V. CONCLUSION

In this paper, we analyze the factors and association rules of traffic accident from 2015 to 2016 in Korea. The analysis shows that most of the accidents are caused by the driver's failure to safely drive. If improving the driver's careless driving attitude, it would be greatly reduce traffic accident. In addition, the result of analysis by the type of accident shows that many large-scale accident occurs by vans. Since the number of passengers are more than other vehicles, more attention of drivers and passengers of van is needed. In the case of medium-scale accident, many accidents occurred between cars, and in the case of small-scale accident, many accidents occurred between car and pedestrians. As for the future work, we would try to predict the severity of traffic accidents by using the detected association rules.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.2016-0-00406, SIAT CCTV Cloud Platform).

REFERENCES

[1] J. Abellán, G. López and J. Oña, "Analysis of traffic accident severity using Decision Rules via Decision Trees," Expert Systems with Applications, vol.40, no.15, pp.6047-6054, Nov. 2013.

[2] U. Castro and Y. Kim, "Data mining on road safety: factor assessment on vehicle accidents using classification models," International Journal of Crashworthiness, vol.21, no.2, pp.104-111, 2014.

[3] M. Chong, A. Abraham and M. Paprzycki, "Traffic Accident Analysis Using Machine Learning Paradigms," Information, vol.29, pp.89-98, 2005.

[4] J.Y. Lee, J.H. Chung and B. Son, "Analysis of traffic accident size for Korean highway using structural equation models," Accident Analysis and Prevention, vol.40, pp.1955-1963, 2008.

[5] L. Li, S. Sherestha and G. Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques," in Proceeding of the Software Engineering Research, Management and Applications, London, July 2017.

[6] "J. Oña, R. Mujalli and F.J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," Accident Analysis and Prevention, vol.43, no.1, pp.402-411, Jan. 2011.

[7] V. Rovšek, M. Batista and B. Bogunović "Identifying the key risk factors of traffic accident injury severity on Slovenian roads using a non-parametric classification tree," Transport, vol.32, no.3, pp.272-281, 2014.

[8] Md.S. Satu, S. Ahamed, F. Hossain, T. Akter and D.Md. Farid, "Mining traffic accident data of N5 national highway in Bangladesh employing decision trees," in Proceeding of Humanitarian Technology Conference, Dhaka, Dec 2017.

[9] S. Sohn and H. Shin, "Pattern recognition for road traffic accident severity in Korea," Ergonomics, vol.44, no.1, pp.107-117, 2010.

[10] M. Taamneh, S. Alkheder and S. Taamneh, "Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates," Journal of Transportation Safety & Security, vol.9, no.2, pp.146-166, 2016.